Measuring Student Growth for Teachers in Non-Tested Grades and Subjects: A Primer

(Please see the Glossary on page 8 for definitions of key terms.)

Focus Question: What options do state education agencies (SEAs) and/or local education agencies (LEAs) have when taking into account student growth in non-tested grades and subjects as part of a rigorous, transparent, and fair system for evaluating teachers and principals?

Race to the Top (RTT) asks that states "design and implement rigorous, transparent, and fair evaluation systems for teachers and principals that differentiate effectiveness using multiple rating categories that take into account data on student growth…as a significant factor." In its definition of student growth, RTT makes the distinction between "tested grades and subjects" and "non-tested grades and subjects."

The Department of Education (ED) defines "tested grades and subjects" as those covered by "the state's assessment under the ESEA" and "non-tested grades and subjects" as those without such data. Because the definition of student growth requires "individual student achievement data from two or more points in time," this definition typically limits the tested grades and subjects to Grades 4–10 in the subjects of English language arts and mathematics. Tested grades and subjects afford relatively large and robust data sets that can be used to measure changes in students' academic achievement. For teachers who teach grades and subjects that do not require a test under ESEA, other measures will need to be used or developed. Since large portions of the teaching workforce in all states work in non-tested grades and subjects, the identification and/or development of additional measures is critical for fulfilling commitments made under RTT proposals. These measures have the potential to inform instruction, build stakeholder commitment, provide a critical dimension to the assessment of teacher effectiveness, and, most important, improve student performance across a broader set of expectations.

RTT guidance on measuring student achievement in non-tested grades and subjects permits "alternative measures of student learning and performance such as student scores on pre-tests and end-of-course tests; student performance on English language proficiency assessments, and other measures of student achievement" if they are "rigorous and comparable across classrooms."

Approaches to the Challenge. There are three general approaches emerging in response to the challenge, though additional approaches are under investigation. It is important to note that these approaches are not mutually exclusive. It is likely that states and districts may want to use a variety of approaches to measuring student growth depending on the assessments available, the cost and benefits of each approach, and the contextual needs within the state.

- **Student Learning Objectives (SLOs):** a participatory method of setting measurable goals, or objectives, based on the specific assignment or class, such as the students taught, the subject matter taught, the baseline performance of the students, and the measurable gain in student performance during the course of instruction. SLOs can be based on the Elementary and Secondary Education Act (ESEA) or other standardized assessments, but they also may be based on teacher-developed or other classroom assessments if they are "rigorous and comparable across classrooms." The general method of SLOs draws on both effective pedagogical practices and approaches to goal setting, and evaluation and task motivation found in multiple professions. In some instances, student learning objectives are shared by a team of job-alike teachers.

- **Other Assessments:** the development and or adaptation of other measures of student growth for non-tested grades and subjects used across schools or districts. These measures may include early

This document was created for the Race to the Top Technical Assistance Network under contract

reading measures; standardized end-of-course assessments; formative assessments; benchmark, interim, or unit assessments; and standardized measures of English language proficiency. Other assessments may be developed at either the SEA or LEA level. It also is possible to imagine teacher-developed assessments of student learning or growth falling into this category when those assessments meet expectations for rigor and comparability across classrooms in a district or across classrooms statewide.

- **Measures of Collective Performance:** the use of measures required by ESEA and/or other standardized assessments used to measure the performance of groups of teachers. Measures of collective performance may assess the performance of the school, grade level, instructional department, teams or other groups of teachers. These measures can take a variety of forms including school-wide student growth measures, team-based collaborative achievement projects, and shared value-added scores for co-teaching situations.

Each approach offers different advantages to SEAs and LEAs for incorporation into teacher and principal evaluation systems. (See Table 1 for further details of each approach and Table 2 for examples of each approach.) Questions states may wish to consider in examining these approaches may include those identified in the following **five categories**:

1. **POLICY REQUIREMENTS.** What statutes, regulations or other factors would impact the design or implementation of assessments used to measure student growth in non-tested subjects? Consider comparability –
    - **a.** How can states assure that student learning is being measured consistently across classrooms in a district or across classrooms statewide?
    - **b.** What steps in development and administration do states or districts need to take to ensure that scores will have the same meaning within the subject area, and that student growth will have a similar interpretation across subjects?

2. **TIMELINES.** What timelines are in place in statute, regulation or elsewhere that would affect the development and implementation of assessments?

3. **ASSESSMENT OPTIONS.** What clearly defined options does the state have for measuring student learning? Consider –
    - **a.** Capacity: What capacity is needed to develop and implement the assessment now and over time, such as district or state data capacity, professional development, or time to administer?
    - **b.** Growth over time**:** How does the approach allow for the measurement of student progress over two or more points in time?
    - **c.** Rigor: How will the state assure that the approach is rigorous?
        - **i.** How does the measure cover expected knowledge and skill appropriately, in terms of the content of questions or tasks included and the coverage of the subject area? Possibilities include expert and stakeholder judgment, quantitative analysis of results, and pilot testing with student feedback.
        - **ii.** What assurances are in place to make sure some students are not placed at a disadvantage by the specific questions or tasks included?
        - **iii.** Does the assessment accurately measure meaningful changes in student achievement in the subject area, either in strictly comparative terms (e.g., some students learned more than other students) or growth toward a standard (e.g., some students made more progress than others toward a goal that will help them be successful)?

This document was created for the Race to the Top Technical Assistance Network under contract

4. **ACTIONS DRIVEN BY ASSESSMENT INFORMATION.** What actions will be informed by data developed from the assessments (such as such as bonus allocations, teacher preparation program approval or teacher tenure decisions)? Consider –
    a. How can information from this model be aggregated across teachers or classrooms to help evaluate principals or other school leaders?

5. **DATA USE PARAMETERS.** How can parameters for data use – such as student demographic variables, assessment methods or procedures for setting student learning objectives – be adjusted to build: a) tighter alignment between policy requirements and actions driven by data from the assessments; and b) greater confidence among practitioners and other stakeholders? Consider –
    a. How easy is the model to explain and describe to stakeholders, such as teachers, leaders, students and parents?
    b. How easily can data from the model be used along with other data to assess teacher or school leader practice?

This document was created for the Race to the Top Technical Assistance Network under contract

Table 1. Details for Each Approach

Approach	Examples in the Field	Advantages	Implementation Challenges
Student Learning Objectives	Student Growth Objectives (Denver, CO) Student Learning Objectives (Charlotte-Mecklenburg, NC) Student Learning Objectives (Austin, TX) Learning Teams Protocol (Los Angeles, CA)	• Adaptable to the wide variety of teaching assignments. • Can be adapted to new assessment structures as they are developed. • Can have credibility with current educators because they are immediately relevant to setting and measuring classroom expectations. • Can have face validity as teacher are often tasked with developing the objectives. • Permit individual incentives, but can also be used in conjunction with measures of school or group performance to create collective incentives. • Permit high degrees of specialization for teachers and students.	• Difficult to create comparability and rigor without common assessments, or common requirements for assessment. • Predictive validation of SLOs in alignment with growth measured by a value-added or student growth measure has been completed only on a limited scale. • Requires significant time and attention from administrators and evaluators.
Other Assessments	New York State Regents Exams Georgia CLASS Keys Using Student Growth to Evaluate Teacher Effectiveness in Delaware (in development) Empowering Effective Teachers (Hillsborough County, FL)	• Create comparability within tested fields of study, and can create similar rigor through multiple classrooms and schools. • A relatively small number can be developed for most commonly taken courses (e.g., graduation required courses). • Permit individual incentives but can also be used in conjunction with measures of school or group performance to create collective incentives. • May be ready to be adapted in some fields, such as current interim or benchmark assessments, and Advanced Placement (AP) or International Baccalaureate (IB) assessments. • Can increase teacher buy-in and professional growth as they can play a critical role in developing tests.	• Assessments will not cover all teaching assignments or courses taken by students. • Current assessments—such as current benchmark or interim assessments, AP or IB assessments—may be designed for purposes other than assessing student growth and/or as a measure of teacher effectiveness. • Assessments may require time for teachers to work together to develop consistent scoring patterns. • Requires attention to ensuring comparability across classrooms.

This document was created for the Race to the Top Technical Assistance Network under contract #ED-ESE-10-O-0087 held by ICF International.

Approach	Examples in the Field	Advantages	Implementation Challenges
Measures of Collective Performance	TAP: The System for Teacher and Student Advancement (National Institute for Excellence in Teaching) Battelle for Kids IMPACT: The DCPS Effectiveness Assessment System for School-Based Personnel (District of Columbia)	• Address the variety of teaching assignments by using agreed upon measures for which schools or groups of teachers share responsibility. • Build collective schoolwide or team-based effort around student achievement.	• Measures of collective performance mask high and low performers in the group and give little information about how individual teachers are doing with their classrooms. • May be perceived as unfair because teachers are held to a measure which they may have had limited ability to impact.

This document was created for the Race to the Top Technical Assistance Network under contract #ED-ESE-10-O-0087 held by ICF International.

Table 2. Example Summary for Each Approach

Approach	Example Summary
Student Learning Objectives	The Austin Independent School District (AISD) REACH program is a pilot strategic compensation program that uses individual teacher-developed, principal-approved student learning objectives (SLOs) as one of four ways to define teacher effectiveness. According to AISD, the stated purpose of using SLOs is to "encourage each teacher to engage in strategic planning and data analysis for their students' learning. SLOs also promote coordinating the goals of teachers with the aims of the whole campus. The end target is increased student learning." Teachers' performance is assessed in REACH using Student Learning Objectives (SLOs). Following AISD guidelines, each teacher creates two SLOs, which must be approved by his or her principal. One SLO must be achieved by 75 percent of the teachers' students in a course. The other SLO can be achieved by a targeted subgroup of students or again by 75 percent of the whole group. To measure the success of each SLO, students are assessed at the beginning and end of the year on available district exams or tests developed by groups of teachers under strict guidelines (including principal approval). Teachers must indicate how each SLO is based on students' needs, aligned with state or national standards, as well as his or her Campus Improvement Plan. Every SLO must be approved by the participant's building principal as well as by the central SLO Core Team. The SLO Team uses a rubric that assesses the rigor of each SLO. Participants submit their SLO to an on-line data management system to which principals and the central SLO Team has access. At the end of the school year, teachers must provide evidence that their students met the SLOs; principals (monitored by AISD staff) determine whether teachers have achieved one or both SLOs. All teachers (core content, Special Education teachers, teachers of ELLs, as well as specialty areas like music, art, physical education, etc.) are part of this process. In 2010, librarians, instructional coaches/specialists, and assistant principals were also required to set SLOs. Teachers choose the outcomes on which they will be judged. This helps ensure that the measure is credible, relevant to the content of the curriculum the teacher intends to teach, and based on broadly used assessments of student learning. SLOs tend to be low-inference measures, so once chosen, there is little question whether or not students met the objective based on the data collected.
Other Standardized Assessments	Hillsborough County Public Schools currently uses a system of pre- and post-tests in each grade and subject to determine merit awards. Specifically, the Florida Comprehensive Assessment Test (FCAT) is used for grades and subjects that take this state exam, but for other grades and core subjects various national norm-referenced tests and district-created tests are used. When the pre/post measures were initially designed for every class, there was some concern that certain specialty tests were not very rigorous and that tests in different subject areas could not be comparable. But over the years, these tests improved and concerns subsided, largely due to the district's involvement of teachers from the full diversity of subject areas. In the case of non-core academic subjects a combination of these district-developed pre/post tests and FCAT reading and math scores are used.

This document was created for the Race to the Top Technical Assistance Network under contract #ED-ESE-10-O-0087 held by ICF International.

Approach	Example Summary
Measures of Collective Performance	The TAP System for Teacher and Student Advancement addresses teacher effectiveness in four ways: multiple career paths, performance-based compensation, ongoing applied professional development, and instructionally focused accountability. All TAP teachers receive a performance assessment score that is based on a Skills, Knowledge, and Responsibilities rubric, classroom achievement gains, and school-level student achievement gains. Classroom and school-level gains are measured using value-added models. Twenty percent of a teacher's summative evaluation is based on school-wide achievement growth. For teachers without classroom achievement gains data, school-wide data is typically used in its place, so for these teachers school-wide achievement growth accounts for as much as 50 percent of a teacher's score.
	Battelle for Kids addresses teacher effectiveness with teachers in co-teaching situations (e.g., general and special education teachers who share students) to develop specific value added scores for teachers based on assignment. Teachers themselves determine their level of attribution by indicating the percentage of time they allocate to each student. This critical process creates teacher buy-in and data accuracy. Teachers working in a true co-teaching situation may indicate equal percentages of attribution for each student (50% and 50%). Resource teachers that work collaboratively with general education teachers may split the student's value added score (e.g., 40% resource room teacher/60% general education classroom teacher).

This document was created for the Race to the Top Technical Assistance Network under contract #ED-ESE-10-O-0087 held by ICF International.

Glossary

The list of terms included here is not meant to be comprehensive but rather to reflect key concepts covered in the primer.

- **Formative assessments.** These are more frequent assessments that target specific standards. They can be used to pre-assess, monitor, and post-assess an individual student's performance relative to specific standards. These assessments may be administered consistently across grade-level content areas, and the resulting data can be used for providing additional instructional supports and/or interventions to assist all students in mastering the targeted and assessed standard.

- **Interim assessments.** Assessments given at regular and specified intervals throughout the school year, designed to evaluate student knowledge and skills relative to a specific set of academic standards. They produce results that can be aggregated (e.g. by course, grade level, or LEA) in order to inform teachers and administrators at the student, classroom, school and LEA level.

- **"Rigorous and comparable across classrooms."** Race to the Top FAQs stated that "measures must be rigorous *(i.e.,* statistically rigorous) and comparable across classrooms in a district or across classrooms statewide. It is not acceptable to use measures of student growth that are only comparable across students within a class."

- **Student achievement.** The definition in Race to the Top application guidance states that student achievement is: (a) for tested grades and subjects: (1) a student's score on the state's assessment under the ESEA, and, as appropriate, (2) other measures of student learning, such as those described in paragraph (b) of this definition, provided they are rigorous and comparable across classrooms; (b) for non-tested subjects: alternative measures of student learning and performance such as student scores on pre-tests and end-of-course tests; student performance on English language proficiency assessments, and other measures of student achievement that are rigorous and comparable across classrooms.

- **Student growth.** The change in student achievement for an individual student between two or more points in time.

This document was created for the Race to the Top Technical Assistance Network under contract #ED-ESE-10-O-0087 held by ICF International.

Leading Sources of Information

Center for Collaborative Education (2010). *Including performance assessments in accountability systems: A review of scale-up efforts.* Boston, MA: Author. Retrieved November 11, 2010, from http://www.ccebos.org/Performance_Assessment_Review_1.10.pdf

National Comprehensive Center for Teacher Quality. (2010). *Challenges in evaluating special education teachers and English language learner specialists.* Washington, DC: Author. Retrieved November 11, 2010, from http://http//www.tqsource.org/publications/July2010Brief.pdf

Prince, C.D., Schuermann, P.J., Guthrie, J.W., Witham, P.J., Milanowski, A.T., & Thorn, C.A. (2009). *The other 69 percent: Fairly rewarding the performance of teachers of non-tested grades and subjects* (A Guide to Implementation: Resources for Applied Practice). Washington, DC: Center for Educator Compensation Reform. Retrieved November 11, 2010, from http://cecr.ed.gov/pdfs/guide/other69Percent.pdf.

Slotnik, W. J., & Smith, M.D. (2004). Catalyst for change: Pay for Performance in Denver (Final Report). Boston: Community Training and Assistance Center. Retrieved November 11, 2010, from http://www.ctacusa.com/PDFs/Rpt-CatalystChangeFull-2004.pdf

Descriptive Information Regarding SLOs

Charlotte Mecklenburg Student Learning Objectives
Guide: http://www.ctacusa.com/PDFs/CMSSLOGuide-2008.pdf
SLO Brief: http://www.ctacusa.com/PDFs/Rpt-StudentLearningObjectives-2008.pdf
Audit: http://www.ctacusa.com/PDFs/Rpt-FocusOnLiteracy-2008.pdf
Implementation Guide for TIF-LEAP: 2010-2011
The link between learning and SLOs: http://www.ctacusa.com/PDFs/Rpt-TyingtoLearning-2008.pdf

Austin Independent School District REACH
SLO Guide:
http://www.austin.isd.tenet.edu/inside/initiatives/compensation/docs/SCI_SLO_Guide_2009-2010.pdf
Austin's Year Two Program Evaluation: http://www.austinisd.org/inside/docs/ope_08-97_AISD_Reach_Year2_Evaluation_ReportII.pdf

Descriptive Information on Benchmark Assessments and Common Formative Assessments

Common Formative Assessments – How to Connect Standards-Based Instruction and Assessment by Larry Ainsworth and Donald Viegut.

Formative Assessment – Making It Happen in the Classroom by Margaret Heritage.

This document was created for the Race to the Top Technical Assistance Network under contract #ED-ESE-10-O-0087 held by ICF International.